

(17.12.04)

日本国特許庁
JAPAN PATENT OFFICE

REC'D 17 DEC 2004

WIPO PCT

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出願年月日 Date of Application: 2003年12月 5日

出願番号 Application Number: 特願2003-406669

[ST. 10/C]: [JP 2003-406669]

出願人 Applicant(s): 株式会社ケンウッド

**PRIORITY
DOCUMENT**
SUBMITTED OR TRANSMITTED IN
COMPLIANCE WITH RULE 17.1(a) OR (b)

2004年12月 2日

特許庁長官
Commissioner,
Japan Patent Office

小川

洋

【書類名】 特許願
【整理番号】 P07-975487
【提出日】 平成15年12月 5日
【あて先】 特許庁長官 殿
【国際特許分類】 G05B 19/4155
G06F 17/20
G10L 15/00

【発明者】
【住所又は居所】 東京都八王子市石川町 2967-3 株式会社ケンウッド内
【氏名】 久米 利佳

【特許出願人】
【識別番号】 000003595
【氏名又は名称】 株式会社ケンウッド

【代理人】
【識別番号】 100095407
【弁理士】
【氏名又は名称】 木村 満

【手数料の表示】
【予納台帳番号】 038380
【納付金額】 21,000円

【提出物件の目録】
【物件名】 特許請求の範囲 1
【物件名】 明細書 1
【物件名】 図面 1
【物件名】 要約書 1
【包括委任状番号】 9903184

【書類名】特許請求の範囲**【請求項1】**

音声を表す音声データを取得し、当該音声データに音声認識を施すことにより、当該音声が表す語句の候補を特定する音声認識手段と、

前記音声認識手段が特定した候補に基づき、前記音声の発話者が欲する当該オーディオ機器の利用態様と、当該利用態様を実現するために当該オーディオ機器に加えるべき操作とを特定し、特定した操作を加えるため前記オーディオ機器を制御するオーディオ機器制御手段と、を備える、

ことを特徴とするオーディオ機器制御装置。

【請求項2】

音声を表す音声データを取得し、当該音声データに音声認識を施すことにより、当該音声が表す語句の候補を特定する音声認識手段と、

前記音声認識手段が特定した候補と、外部のオーディオ機器が使用される環境の状況を示す環境データ及び／又は当該オーディオ機器の動作状態を示す動作状態データとにに基づき、前記音声の発話者が欲する当該オーディオ機器の利用態様と、当該利用態様を実現するために当該オーディオ機器に加えるべき操作とを特定し、特定した操作を加えるため前記オーディオ機器を制御するオーディオ機器制御手段と、を備える、

ことを特徴とするオーディオ機器制御装置。

【請求項3】

音声を表す音声データを取得し、当該音声データに音声認識を施すことにより、当該音声が表す語句の候補を特定する音声認識手段と、

前記音声データに基づき、前記音声の発話者又は発話者の属性を特定する発話者特定手段と、

前記音声認識手段が特定した候補と、前記発話者特定手段が特定した発話者又は発話者の属性と、外部のオーディオ機器が使用される環境の状況を示す環境データ及び／又は当該オーディオ機器の動作状態を示す動作状態データとにに基づき、前記音声の発話者が欲する当該オーディオ機器の利用態様と、当該利用態様を実現するために当該オーディオ機器に加えるべき操作とを特定し、特定した操作を加えるため前記オーディオ機器を制御するオーディオ機器制御手段と、を備える、

ことを特徴とするオーディオ機器制御装置。

【請求項4】

前記環境データは、前記オーディオ機器が使用される環境の現在位置を示すデータからなっている、

ことを特徴とする請求項2又は3に記載のオーディオ機器制御装置。

【請求項5】

前記環境データは、前記オーディオ機器が使用される環境の遮音状況を示すデータからなっている、

ことを特徴とする請求項2、3又は4に記載のオーディオ機器制御装置。

【請求項6】

音声を表す音声データを取得し、当該音声データに音声認識を施すことにより、当該音声が表す語句の候補を特定する音声認識ステップと、

前記音声認識ステップで特定された候補に基づき、前記音声の発話者が欲する当該オーディオ機器の利用態様と、当該利用態様を実現するために当該オーディオ機器に加えるべき操作とを特定し、特定した操作を加えるため前記オーディオ機器を制御するオーディオ機器制御ステップと、より構成される、

ことを特徴とするオーディオ機器制御方法。

【請求項7】

音声を表す音声データを取得し、当該音声データに音声認識を施すことにより、当該音声が表す語句の候補を特定する音声認識ステップと、

前記音声認識ステップで特定された候補と、外部のオーディオ機器が使用される環境の

状況を示す環境データ及び／又は当該オーディオ機器の動作状態を示す動作状態データとに基づき、前記音声の発話者が欲する当該オーディオ機器の利用態様と、当該利用態様を実現するために当該オーディオ機器に加えるべき操作とを特定し、特定した操作を加えるため前記オーディオ機器を制御するオーディオ機器制御ステップと、より構成される、ことを特徴とするオーディオ機器制御方法。

【請求項8】

音声を表す音声データを取得し、当該音声データに音声認識を施すことにより、当該音声が表す語句の候補を特定する音声認識ステップと、

前記音声データに基づき、前記音声の発話者又は発話者の属性を特定する発話者特定ステップと、

前記音声認識ステップ特定された候補と、前記発話者特定ステップで特定された発話者又は発話者の属性と、外部のオーディオ機器が使用される環境の状況を示す環境データ及び／又は当該オーディオ機器の動作状態を示す動作状態データとに基づき、前記音声の発話者が欲する当該オーディオ機器の利用態様と、当該利用態様を実現するために当該オーディオ機器に加えるべき操作とを特定し、特定した操作を加えるため前記オーディオ機器を制御するオーディオ機器制御ステップと、より構成される、

ことを特徴とするオーディオ機器制御方法。

【書類名】明細書

【発明の名称】オーディオ機器制御装置、オーディオ機器制御方法及びプログラム

【技術分野】

【0001】

この発明は、オーディオ機器制御装置、オーディオ機器制御方法、データ出力装置及びプログラムに関する。

【背景技術】

【0002】

近年、音声認識の技術を用いて音声を認識し、認識結果に応答して電気機器などを制御する手法が用いられている。この手法は、具体的には、入力した音声が表す単語を識別し、識別された単語が所定のキーワードに合致するか否かを判別して、判別結果に基づいて外部の機器を制御するものである。このような手法は、例えばオーディオ機器の制御にも用いられている（例えば、特許文献1及び特許文献2参照）。

【特許文献1】特開平4-324312号公報

【特許文献2】特願2000-383809号公報

【発明の開示】

【発明が解決しようとする課題】

【0003】

しかし、人間が言語の形で発する指示を完全に認識するのは困難である。このため、上述の手法では、人間が言語の形で発する指示に適切に応答することができない場合があった。

【0004】

この発明は上記実状に鑑みてなされたものであり、人間が言語の形で発する指示に適切に応答して機器を制御できるオーディオ機器制御装置、オーディオ機器制御方法及びプログラムを提供することを目的とする。

【課題を解決するための手段】

【0005】

上記目的を達成するため、この発明の第1の観点にかかるオーディオ機器制御装置は、音声を表す音声データを取得し、当該音声データに音声認識を施すことにより、当該音声が表す語句の候補を特定する音声認識手段と、

前記音声認識手段が特定した候補に基づき、前記音声の発話者が欲する当該オーディオ機器の利用態様と、当該利用態様を実現するために当該オーディオ機器に加えるべき操作とを特定し、特定した操作を加えるため前記オーディオ機器を制御するオーディオ機器制御手段と、を備える、

ことを特徴とする。

【0006】

また、この発明の第2の観点にかかるオーディオ機器制御装置は、音声を表す音声データを取得し、当該音声データに音声認識を施すことにより、当該音声が表す語句の候補を特定する音声認識手段と、

前記音声認識手段が特定した候補と、外部のオーディオ機器が使用される環境の状況を示す環境データ及び／又は当該オーディオ機器の動作状態を示す動作状態データとに基づき、前記音声の発話者が欲する当該オーディオ機器の利用態様と、当該利用態様を実現するために当該オーディオ機器に加えるべき操作とを特定し、特定した操作を加えるため前記オーディオ機器を制御するオーディオ機器制御手段と、を備える、

ことを特徴とする。

【0007】

また、この発明の第3の観点にかかるオーディオ機器制御装置は、音声を表す音声データを取得し、当該音声データに音声認識を施すことにより、当該音声が表す語句の候補を特定する音声認識手段と、

前記音声データに基づき、前記音声の発話者又は発話者の属性を特定する発話者特定手

段と、

前記音声認識手段が特定した候補と、前記発話者特定手段が特定した発話者又は発話者の属性と、外部のオーディオ機器が使用される環境の状況を示す環境データ及び／又は当該オーディオ機器の動作状態を示す動作状態データとに基づき、前記音声の発話者が欲する当該オーディオ機器の利用態様と、当該利用態様を実現するために当該オーディオ機器に加えるべき操作とを特定し、特定した操作を加えるため前記オーディオ機器を制御するオーディオ機器制御手段と、を備える、

ことを特徴とする。

【0008】

前記環境データは、例えば、前記オーディオ機器が使用される環境の現在位置を示すデータからなっている。

【0009】

前記環境データは、例えば、前記オーディオ機器が使用される環境の遮音状況を示すデータからなっている。

【0010】

また、この発明の第4の観点にかかるオーディオ機器制御方法は、

音声を表す音声データを取得し、当該音声データに音声認識を施すことにより、当該音声が表す語句の候補を特定する音声認識ステップと、

前記音声認識ステップで特定された候補に基づき、前記音声の発話者が欲する当該オーディオ機器の利用態様と、当該利用態様を実現するために当該オーディオ機器に加えるべき操作とを特定し、特定した操作を加えるため前記オーディオ機器を制御するオーディオ機器制御ステップと、より構成される、

ことを特徴とする。

【0011】

また、この発明の第5の観点にかかるオーディオ機器制御方法は、

音声を表す音声データを取得し、当該音声データに音声認識を施すことにより、当該音声が表す語句の候補を特定する音声認識ステップと、

前記音声認識ステップで特定された候補と、外部のオーディオ機器が使用される環境の状況を示す環境データ及び／又は当該オーディオ機器の動作状態を示す動作状態データとに基づき、前記音声の発話者が欲する当該オーディオ機器の利用態様と、当該利用態様を実現するために当該オーディオ機器に加えるべき操作とを特定し、特定した操作を加えるため前記オーディオ機器を制御するオーディオ機器制御ステップと、より構成される、

ことを特徴とする。

【0012】

また、この発明の第6の観点にかかるオーディオ機器制御方法は、

音声を表す音声データを取得し、当該音声データに音声認識を施すことにより、当該音声が表す語句の候補を特定する音声認識ステップと、

前記音声データに基づき、前記音声の発話者又は発話者の属性を特定する発話者特定ステップと、

前記音声認識ステップ特定された候補と、前記発話者特定ステップで特定された発話者又は発話者の属性と、外部のオーディオ機器が使用される環境の状況を示す環境データ及び／又は当該オーディオ機器の動作状態を示す動作状態データとに基づき、前記音声の発話者が欲する当該オーディオ機器の利用態様と、当該利用態様を実現するために当該オーディオ機器に加えるべき操作とを特定し、特定した操作を加えるため前記オーディオ機器を制御するオーディオ機器制御ステップと、より構成される、

ことを特徴とする。

【発明の効果】

【0013】

この発明によれば、人間が言語の形で発する指示に適切に応答して機器を制御できるオーディオ機器制御装置、オーディオ機器制御方法及びプログラムが実現される。

【発明を実施するための最良の形態】

【0014】

以下、図面を参照して、この発明の実施の形態を、車両内に搭載されて用いられるカーオーディオシステムを例として説明する。

図1は、このカーオーディオシステムの構成を示すブロック図である。図示するように、このカーオーディオシステムは、音声入力部1と、言語解析部2と、音声合成処理部3と、音声出力部4と、入出力対象機器群5と、エージェント処理部6により構成されている。

【0015】

音声入力部1は、例えば、マイクロフォン、A F (Audio Frequency) 増幅器、サンプラー及びA／D (Analog-to-Digital) コンバータなどにより構成されている。音声入力部1は、自己のマイクロフォンが集音した音声を表す音声信号を増幅し、サンプリングしてA／D変換することにより、当該音声を表すデジタル形式の音声データを生成する。そして、この音声データを言語解析部2へと供給する。

【0016】

言語解析部2、音声合成処理部3及びエージェント処理部6は、それぞれ、例えば、C P U (Central Processing Unit) 等からなるプロセッサと、このプロセッサが実行するプログラムを記憶するハードディスク装置等の不揮発性メモリと、このプロセッサのワークエリアとなる記憶領域を有するR A M (Random Access Memory) 等の揮発性メモリにより構成されている。

なお、言語解析部2、音声合成処理部3及びエージェント処理部6の一部又は全部の機能を单一のプロセッサや单一の不揮発性メモリや单一の揮発性メモリが行うようにしてもよい。

【0017】

言語解析部2は、音声入力部1より供給された音声データに音声認識を施す処理を行うことにより、この音声データが表している可能性のある単語の候補と、この候補の尤度(スコア)とを特定する。音声認識の手法は任意であり、また、候補は複数特定されてよい。そして、特定した候補及び当該候補のスコアを示すデータ(以下、単語データと呼ぶ)を生成し、エージェント処理部6へと供給する。

【0018】

音声合成処理部3の不揮発性メモリは、更に、単語の波形を表すデータを記憶する音片合成用データベースと、音素の波形を表すデータを記憶する素片合成用データベースとを記憶する。そして、音声合成処理部3は、音片合成用データベースや素片合成用データベースに格納されているデータを用いて、録音編集方式や規則合成方式の手法により、エージェント処理部6より供給された文章データを読み上げる音声を表すデジタル形式の音声データを生成する。そして、生成した音声データを音声出力部4に供給する。

【0019】

音声出力部4は、例えば、D／A (Digital-to-Analog) コンバータ、A F 増幅器及びスピーカなどにより構成されている。音声出力部4は、音声合成処理部3より供給された音声データをD／A変換して増幅し、得られたアナログ信号を用いてスピーカを駆動することにより、当該音声データが表す音声を再生する。

【0020】

入出力対象機器群5には、例えば、C D (Compact Disc) プレーヤー51や、M D (Mi ni Disc) プレーヤー52や、車両位置特定部53や、窓開閉制御部54や、クロック55などが含まれている。

【0021】

C D プレーヤー51は、自己に装着されたC Dに記録された音声データが表す音声の再生や再生の停止、ミュート、早送り再生、巻き戻し再生、一時停止C Dの排出、等の動作を、エージェント処理部6が供給する制御信号に従って行う。

M D プレーヤー52は、自己に装着されたM Dに記録された音声データが表す音声の再

生や再生の停止、ミュート、早送り再生、巻き戻し再生、一時停止、CDの排出、等の動作を、エージェント処理部6が供給する制御信号に従って行う。

【0022】

車両位置特定部53は、GPS (Global Positioning System) の移動局を用いた公知のカーナビゲーションシステムなどより構成されており、エージェント処理部6が供給する制御信号に従って車両の現在位置を検知し、検知した現在位置が地図上でどの位置に当たるかを表すデータを生成して、エージェント処理部6に供給する。

【0023】

窓閉鎖制御部54は、モーターや、制御信号に従ってモータの回転及び停止を制御する制御回路や、モーターの回転に従って車両の窓ガラスを動かすためのウインチ等から構成されており、自己に供給される制御信号に従って、車両の窓の開閉を行う。また、窓閉鎖制御部54の制御回路は、窓閉鎖制御部54の動作状態を表すデータとして、例えば、開閉する対象の窓が開いている量を示すデータを生成して出力する。

【0024】

クロック55は、例えば水晶発振器やカウンタ回路等より構成されており、現在の時刻を示すデータを連続的に生成し、エージェント処理部6へと供給する。尚、カーナビゲーションシステムを介してGPSから得られる時刻情報を用いてもよい。

【0025】

エージェント処理部6の説明に戻ると、エージェント処理部6は更に、モデムやパケット通信端末等からなる通信制御装置を備え、この通信制御装置を介して外部のネットワークに接続している。

【0026】

また、エージェント処理部6の不揮発性メモリは、更に、単語データベースを記憶する。単語データベースは、単語を示すデータと、この単語がどのような概念の下にグルーピングされているかを示すための単語グルーピング用のフラグ1個以上と、互いに対応付けた形で格納するデータベースである。

【0027】

1個の単語に対応付けられている各フラグは、互いに異なる概念に対応付けられている。そして、フラグが所定の値（以下では、この値は“1”であるとする）を示す場合は、このフラグに対応付けられた単語が、このフラグに対応付けられた概念の下にグルーピングされていることを示す。一方、このフラグが他の値（例えば“0”）を示す場合は、この単語がこの概念の下にはグルーピングされていないことを示す。

【0028】

図2は、グルーピング用のフラグの具体例を模式的に示す図である。図2に示すように、単語「聞きたい」、「停止」及び「つまらない」に、単語グルーピング用のフラグが3ビットずつ対応付けられているものとする。また、これらのいずれの単語に対応付けられたビット群についても、その最上位のフラグが「CDプレーヤー又はMDプレーヤーの操作」という概念に対応付けられており、上位から2ビット目のフラグが「感想」という概念に対応付けられており、上位から3ビット目のフラグが「否定」という概念に対応付けられているとする。一方、図示するように、単語「聞きたい」に対応付けられている3ビットのフラグ群の値が2進数“110”であり、単語「停止」に対応付けられているフラグ群の値が2進数“101”であり、単語「つまらない」に対応付けられているフラグ群の値が2進数“011”であるとする。

この場合、このフラグ群は、概念「CDプレーヤー又はMDプレーヤーの操作」の下には単語「聞きたい」及び「停止」がグルーピングされており、概念「感想」の下には単語「聞きたい」及び「つまらない」がグルーピングされており、概念「否定」の下には単語「停止」及び「つまらない」がグルーピングされていることを示す。

【0029】

また、エージェント処理部6の不揮発性メモリは、更に、楽曲データベースを記憶する。

楽曲データベースは、楽曲の曲名と、この楽曲を実演するアーティストのアーティスト名と、この楽曲が記録されているCD及び／又はMDと、当該CD及び／又はMDでこの楽曲が記録されているトラックの番号と、を互いに対応付けて格納するデータベースである。

【0030】

エージェント処理部6の不揮発性メモリは、更に、処理項目データベース及びワイヤデータベースを記憶している。

【0031】

処理項目データベースは、エージェント処理部6が行うトリガ取得処理、判別処理及び入出力処理の内容を、処理項目（ポインタ）毎に記述するデータを格納したデータベースである。

【0032】

トリガ取得処理の内容を記述するデータは、一連の処理を開始させるトリガとして取得するデータの内容を指定するデータと、後述する進行方向の確率係数とからなっている。取得するデータは任意のものであってよく、例えば、CDプレーヤー51やMDプレーヤー53がそれぞれ動作中及び停止中のうちどちらの状態にあるかを示すデータや、窓が開いている量を示すデータや、室内の温度を示すデータや、言語解析部2より供給される上述の単語データであればよい。あるいは、エージェント処理部6自身が行う処理から引き渡されるデータであってもよい。また、トリガ取得処理で取得されるデータが単語データである場合は、当該単語データが表す単語に代えて、当該単語データが表す単語がグループ化されている概念が記述されてもよい。ただし、トリガ取得処理の内容は、複数のトリガ取得処理が互いに同一の単語を表す単語データを取得することができないように記述されるものとする。

【0033】

判別処理の内容を記述するデータは、判別条件と、判別結果としてとり得る結果のリストと、後述する戻り方向の確率係数とを、判別処理別に記述したデータを含んでいる。また、判別処理の内容を記述するデータは、進行方向の確率係数を、とり得る判別結果毎に記述したデータを含んでいる。

【0034】

判別処理では、判別に用いるデータを、判別を行うに先立って、この判別に用いるデータを任意の取得源から取得する場合があってもよい。取得源としては、例えば、言語解析部2や、エージェント処理部6が実行する他の処理や、入出力対象機器群5に属する機器や、その他外部の機器などが考えられる。そしてこの場合、判別処理の内容を記述するデータは、例えば、判別に用いるデータの取得源を指定するデータを更に含んでいればよい。

【0035】

また、判別処理では、所定のデータを、判別に先立って所定の出力先に出力するようにしてよい。具体的には、例えば、所定の質問を表すデータを、判別に先立って音声合成処理部3に引き渡す、等が考えられる。判別処理において所定のデータを判別に先立って所定のデータを出力する場合、判別処理の内容を記述するデータは、例えば、出力するデータの内容と、このデータの出力先とを指定するデータを更に含んでいればよい。

【0036】

入出力処理の内容を記述するデータは、入力あるいは出力するデータの内容を指定するデータからなっている。入力ないし出力するデータは任意の内容を有していてよく、例えば、出力するデータは、音声合成処理部3を介して音声出力部4に発生させる音声の読みを表すデータや外部の機器を制御する制御信号であってもよいし、入力するデータは外部の機器から供給されるデータであってもよい。

【0037】

ワイヤデータベースは、複数の処理間の遷移を記述するデータ（以下、このデータをワイヤと呼ぶ）の集合からなっている。ワイヤは、例えば図3に示すような書式で記述され

たデータからなっており、図示するように、先行する処理から後続する処理への遷移について、当該先行の処理と、当該後続の処理と、当該遷移に対して与えられた重み係数と、を指定するデータである。ただし、先行の処理が判別処理である場合は、当該判別処理のどの判別結果からの遷移であるか、まで記述される必要がある。

【0038】

そして、エージェント処理部6は、処理項目データベース及びワイヤデータベースが全体として表しているフローを実行する。

具体的には、例えば、エージェント処理部6は、「再生して」という、楽曲の再生を希望する単語を含む音声が入力されることにより、この単語を表す単語データが言語解析部2より供給されると、この単語データに応答して、楽曲の再生が指示されることを認識し、「MDを再生する？ CDを再生する？」と質問する音声を、音声合成処理部3を介して音声出力部4に再生させ、「CD」又は「MD」の単語を含む音声が入力され、これらの単語のいずれかを表す単語データが言語解析部2より供給されると、供給された単語が示す記録媒体（すなわちCD又はMD）に記録された音声を、CDプレーヤー51又はMDプレーヤー52に指示して再生させる。

【0039】

また「退屈」という概念の下にグルーピングされた単語を表す単語データが言語解析部2より供給された場合も、エージェント処理部6は、楽曲の再生が指示されたことを認識した場合の手順と同様の手順を経て、記録媒体に記録された音声を、CDプレーヤー51又はMDプレーヤー52に指示して再生させる。

【0040】

ただし、エージェント処理部6は、楽曲の再生が指示されたことを認識した場合は、楽曲の再生を指示する処理に移る前に、窓開閉制御部54より供給されるデータに基づいて、車両の窓が開いているか否かを判別する。そして、開いていると判別した場合は、楽曲の再生の指示を行わず、代わりに、車内の遮音状況が悪いことを認識して、窓開閉制御部54に窓を閉めるよう指示するか、あるいは音声合成処理部3を介し、音声出力部4に、窓を閉めるよう促すメッセージを読み上げる音声を発生させる。また、楽曲の再生中に、「ボリュームを上げて」という意味を示す単語データが高いスコアを示す場合、ボリュームの上昇処理に先立って窓の開閉状態を検出し、窓が開いていれば「窓が開いているよ。窓を閉める？」という音声データを音声合成処理部3で形成し、音声出力部4で出力させるようにしてもよい。

【0041】

「停止して」という単語を表す単語データが言語解析部2より供給されると、エージェント処理部6はこの単語データに応答して、CDプレーヤー51及びMDプレーヤー52から動作状態を示すデータを各々取得し、取得した各データに基づいて、CDプレーヤー51又はMDプレーヤー52のどちらが音声の再生中であるかを特定して、特定された方に音声再生の停止を指示し、再生を停止させる。

【0042】

「うるさい」という単語を表す単語データが言語解析部2より供給されると、エージェント処理部6はこの単語データに応答して、CDプレーヤー51及びMDプレーヤー52から各々の動作状態を示すデータを取得し、取得した各データに基づいて、CDプレーヤー51又はMDプレーヤー52のどちらが音声の再生中であるかを特定して、特定された方に音声のミュートを指示して音を消したり、音量を減少させる。

【0043】

曲名又はアーティスト名を表す単語データが言語解析部2より供給されると、エージェント処理部6は楽曲データベースを検索して、この曲名又はアーティスト名に合致する楽曲を特定し、特定された楽曲を記録している記録媒体を更に特定して、この記録媒体に記録されている当該楽曲を、CDプレーヤー51又はMDプレーヤー52に指示して再生させる。楽曲が複数特定された場合は、該当する複数の楽曲すべてを再生させてもよいし、制御対象機器群5から取得する任意のデータに基づいて、任意の基準に従い、再生する対

象の楽曲を選択してもよい。例えば、車両位置特定部53が供給するデータが、このカーオーディオシステムを搭載する車両が高速道を走行中であることを示しているときは、テンポが一定値より速い楽曲を選択する、など。この場合、例えば楽曲データベースが、各楽曲のテンポを示すデータを記憶していればよい。また同様に、例えば、海沿いの道を走行中の場合には、この環境に関連付けられた楽曲が再生される。これは、演奏者等のイメージから関連付けられてもよいし、また、曲のタイトルに「波」「海」などの言葉が含まれている楽曲を選曲するようにしてもよい。

クロック55から得られる時刻情報や季節情報に基づいて、再生する楽曲を提案するようにしてもよい。また、音声認識を行う際に抽出された話者の音声のピッチ情報から、発話者を特定し、発話者の嗜好に合わせた楽曲の提案を行うようにしてもよい。この場合、予め想定される発話者の音声のピッチ情報をデータベースとして管理し、入力された音声信号からピッチ情報のデータベースを検索し、一致する若しくは最も近いピッチを有する発話者を特定する。同様に、発話者毎の嗜好情報をデータベースとして管理する。これは、例えば、特定の発話者が選曲する楽曲をデータベース化し、嗜好を判別するようにすればよい。嗜好の分類は、ジャンルやアーティスト、楽曲の年代等の情報を用いることができる。

【0044】

エージェント処理部6は、クロック55から、所定の条件に合致する日時を示すデータが供給されると、これに応答して、ネットワークを介して外部のサーバ等に接続し、楽曲データベースに追加するデータをダウンロードして、楽曲データベースに追加する。また、エージェント処理部6は、ユーザに告知すべき情報（該当するアーティストに係る商品の販売スケジュールなど）を表すデータをダウンロードして、音声合成処理部3を介し、音声出力部4に、この情報を音声の形で提供させてもよい。また、ユーザが要求したアーティストや楽曲が現在データとして保有されていない場合、外部のサーバ等に接続し、該当するアーティストや楽曲等を楽曲データベースにダウンロードして追加してもよい。尚、この場合、ダウンロードを開始する前に「希望した楽曲が現在保有されていないので、購入してダウンロードするか？」という問い合わせのメッセージを出力するようにしてもよい。

【0045】

また、エージェント処理部6は、アーティスト名を表す単語データが言語解析部2より供給されると、これに応答して、ネットワークを介して外部のショッピングサイトのサーバ等に接続し、該当するアーティストに係る商品の予約手続が可能な状態として、当該商品の予約を促すメッセージを、例えば音声合成処理部3を介し、音声出力部4に、音声の形で発させてもよい。

【0046】

なお、エージェント処理部6は、例えば言語解析部2が単語データを複数供給した場合などにおいては、複数の判別処理を並行して行う。またこの場合、エージェント処理部6は同一の単語を入力の対象とする処理（例えば、トリガ取得処理や、判別処理におけるデータの入力）が複数があって、該当する単語を表す単語データが言語解析部2より供給された場合は、これらの処理すべてを並行して行う。

【0047】

また、エージェント処理部6は、図4にフローを示すように、先行する第1の処理P1を実行して後続の第2の処理P2に遷移するようにワイヤW01により定義されており、また、第2の処理P2を実行して後続の第3の処理P3に遷移するようにワイヤW03によって定義されているとき、以下の処理を行う。（なお、図示するように、処理P1～P3のいずれについても、各進行方向の確率係数はいずれも0.5であるものとする。）

【0048】

まず、エージェント処理部6が第1の処理P1に到達しているとき、エージェント処理部6は、ワイヤW01、W03及びW05のそれぞれの重み係数を計算し、計算結果をワイヤW01、W03及びW05に書き込む。これら値は、各処理に予め設定されている進

行方向の確率係数によって決定される。

【0049】

具体的には、処理P1に処理が到達したとき、ワイヤW01の重み係数は、処理P1のワイヤに係る確率係数の値すなわち0.5となる。

処理P2のワイヤW03の重み係数は、処理P1のワイヤW1に係る確率係数0.5に処理P2のワイヤW03に係る確率係数0.5を乗じた結果すなわち0.25となる。

ワイヤW05の重み係数は、処理P1のワイヤW1に係る確率係数0.5に処理P2のワイヤW03に係る確率係数0.5を乗じた結果に更に処理P3のワイヤW5に係る確率係数0.5を乗じた結果、すなわち0.125となる。

【0050】

このようにして、ある処理を基点としたときのそれぞれのワイヤの重み係数が計算される。よって、現在の状態が遷移すると、現在の処理を基点にしてその都度重み係数が計算されることになる。

【0051】

具体的には、現在の状態が処理P2に遷移すると、ワイヤW03の重み係数は、処理P2のワイヤW03に係る確率係数に等しい値0.5となり、ワイヤW05の重み係数は処理P2のワイヤW03に係る確率係数0.5と処理P3のワイヤW5に係る確率係数0.5との積すなわち0.25となる。またこのとき、エージェント処理部6は、逆方向、つまり処理P1に戻る方向に係るワイヤW01の重み係数も再度書き込まれる。処理P2に遷移した場合では、ワイヤW01に係る戻り方向の確率係数0.1がそのままワイヤW01の重み係数となる。処理P3に遷移した場合は更に、ワイヤW03に係る戻り方向の確率係数0.1がそのままワイヤW03の重み係数となる。そして、処理P3に遷移した状態におけるワイヤW01の重み係数は、処理P3に遷移した状態におけるワイヤW03の重み係数0.1に、処理P2の戻り方向の確率係数0.1を乗じた値すなわち0.01となる。

【0052】

重み係数の計算は、関連するフローの処理のみではなく、全てのフローの全てのワイヤについて設定される。ここで現在の処理に関連のないワイヤについては、予め定められた低い計数値を割り当てるようにすればよい。しかし、特にトリガ取得処理を先行の処理とするワイヤについては、重み係数をある程度高く設定するようとする。こうすることによって、直前までなされていた会話と著しく異なる内容の会話にもジャンプすることが可能になる。

【0053】

そして、エージェント処理部6は、判別条件に係る（1個又は複数個の）単語データが言語解析部2より供給されると、以下の処理を行う。まず、供給された単語データが示す単語の尤度（スコア）と、各ワイヤの重み係数との積を計算する。

例えば、図4に示すフローを実行している場合において、単語「聞きたい」に対するスコアが80%で、単語「窓」に対するスコアが50%であることを示す単語データが入力されたとし、また、単語「聞きたい」を示す単語データを入力する処理を先行の処理とするワイヤW01の重み係数が0.5、単語「窓」を示す単語データを入力する処理を先行の処理とするワイヤW05の重み係数が0.125であったとする。この場合、ワイヤW01及びW05について求められる積は、数式1及び2に示すとおりとなる。

【0054】

(数1) ワイヤW01についての積：「聞きたい」に対するスコア80%×ワイヤW01の重み係数0.5 = 4.0

【0055】

(数2) ワイヤW05についての積：「窓」に対するスコア50%×ワイヤW05の重み係数0.125 = 6.25

【0056】

エージェント処理部6は、スコアと重み係数との積を求める上述の処理を、フローが有

するすべてのワイヤについて行う。その結果、例えばワイヤW1について求めた積が最も高い値を示した場合、入力された単語データは単語「聞きたい」を示すものであったと認識して、ワイヤW01が後続の処理としている処理P2に遷移することになる。

【0057】

なお、トリガ取得処理からの遷移に対しては、ある程度高い重み係数を設定しておくとよい。具体的には、例えば、「窓を開けて」という単語を示す単語データを取得するトリガ取得処理からの遷移に対しては、例えば重み係数0.5を与えておく。そうすると、例えばエージェント処理部6の処理がこのトリガ取得処理に係属している場合において、ユーザーが「窓を開けて」と発音し、例えば単語「窓を開けて」に対するスコアが90%である単語データが得られれば、このスコアと、「ユーザが「窓を開けて」と言ったか否かの判断」に係るワイヤの重み係数との積は、90%×0.5すなわち45となる。この値が他のワイヤとの積の値と相手は通常得られないような大きな値であれば、入力された音声は「窓を開けて」であったと認識され、エージェント処理部6の処理が、このトリガ取得処理に後続する処理へとジャンプする可能性が高くなる。一方で、他のワイヤの重み係数を極めて低く設定しておけば、これら他のワイヤにより定義されている遷移が起こる可能性は極めて低くなり、結果として、ある程度想定される会話の流れに沿って認識率を向上させることができる。

【0058】

この実施の形態では、戻り方向への遷移も起こり得る。しかし、現実的には会話を戻すことは好ましくないことが多い。そこで、戻り方向の確率係数は、進行方向の確率係数に比べて低い値に設定するようにすればよい。そうすると、入力された音声から高いスコアの音声データが仮に得られても、戻り方向の確率係数が重み係数として書き込まれたワイヤについて求めた積は低い値となるため、戻り方向への遷移の可能性を低く抑えることができる。

また、エージェント処理部6は、求めた積の値が所定の条件に合致しないような処理（たとえば、積の値が所定値に達しないような処理）は、遷移を実行する対象から除外するよう取り扱ってもよい。

【0059】

なお、例えば図3に示しているように、ワイヤは、処理項目から処理項目への遷移という形で遷移を定義する。そして、ワイヤを図3に示すような形態で記述してデータベースに格納することにより、各処理項目同士の関係を、あたかもコンピュータのマクロ処理のように定義することが可能になる。これによって、各処理項目を容易に接続することができる。

【0060】

また、トリガとなる処理項目は、実際には接続されるワイヤに係る認識対象単語等（他の入力対象機器群からの入力の場合もあり得る）のスコアの判定になるので、ワイヤにおいてトリガ処理項目はワイヤの開始点として定義されず、ワイヤそのものが遷移元として定義されることになる。

【0061】

更に、上述のように各処理項目の接続関係をワイヤによって定義することによって、簡単にワイヤを追加することができる。例えば、「暑い」という音声入力の後に、ユーザーが休憩することを意図して「ファミリーレストランを探して」という音声を入力する機会が多い場合、ファミリーレストランの検索処理項目に対して自動でワイヤを追加する。そうすると、ワイヤが自動で追加された後には、ファミリーレストラン検索処理項目に接続されたワイヤの重み係数をある程度大きくすることで、当該入力「ファミリーレストランを探して」に適切に対応することができるようになる。（ただしこの場合、エージェント処理部6は、例えばファミリーレストランの位置を示す情報を含んだ地図データ等を記憶し、あるいは外部の地図データ等にアクセスするものとする。）

このワイヤの自動追加は、ある処理項目からある処理項目へのジャンプの回数を計数し、これが所定回数に達したときに自動で行うようにすればよい。

【0062】

以上説明したカーオーディオシステムでは、図4を用いて説明した、ワイヤと処理項目の関係、ワイヤに設定される条件、また、重み係数（確率係数も同様）が適宜適用される。

従って、以上説明したこのカーオーディオシステムは、処理の内容を示すデータやワイヤが適切に記述されれば、制御する対象である機器や加える制御の内容を完全に特定することを必ずしも必要とせずに、ユーザが発した言語に応答し、この言語からユーザの欲求を推測し、この欲求を満たすためにどの機器にどのような制御を加えればよいかを適切に判断して、判断結果に従った制御を機器に加えることができるようになる。

【0063】

なお、このカーオーディオシステムの構成は上述のものに限られない。

例えば、入出力対象機器群5に属する機器は、必ずしも直接にユーザの欲求を満たす結果をもたらす機器である必要はなく、例えば、外部の表示装置等を制御してユーザに特定の行動をとるよう促すメッセージを出力する機器（例えば、液晶ディスプレイ等の表示装置）からなっていてもよい。

【0064】

また、単語データベースは、必ずしも単語を示すデータのみならず、複数の単語からなる語句を示すデータを単語データベースの要素として記憶するようにしてもよいし、単語の一部あるいは音素を示すデータを単語データベースの要素として記憶するようにしてもよい。また、単語等は必ずしも特定の概念の下にグルーピングされている必要はなく、グルーピングを行う場合も、グルーピングを行うために用いられるデータは、必ずしもフラグの集合の形をとっていなくてもよい。

【0065】

また、エージェント処理部6は、ワイヤに記述された重み係数を、過去に当該ワイヤが表す遷移を実行した数などに基づき所定の基準に従って変化させ、重み係数が変化後の値となるようにワイヤを書き換えてよい。

具体的には、例えば、ワイヤデータベースに、それぞれのワイヤについて、当該ワイヤが表す遷移が実行された回数を記憶しておく。そしてエージェント処理部6は、当該遷移が新たに行われる毎に、この回数の値を書き換えることにより、この回数の値を1ずつインクリメントし、それぞれのワイヤに記述された重み係数を、例えば、当該ワイヤについて記憶された回数に比例した値と書き換える。

【0066】

また、エージェント処理部6は、判別処理や入出力処理において出力するデータを、これらの処理に引き渡されたデータや、これらの処理に伴って入力したデータや、その他任意の条件に従って変化させるようにしてもよい。

【0067】

また、このカーオーディオシステムは、エージェント処理部6の制御に従って画像を出力するための表示装置（例えば、液晶ディスプレイ等）を備えていてもよく、エージェント処理部6は、入出力処理や判別処理において、処理毎に所定の画像を表示させるようこの表示装置を制御してもよい。

【0068】

また、エージェント処理部6は、1個の入力処理や1個の判別処理において、連続して発話される等した複数の単語データを一括して取得するようにしてもよい。また、エージェント処理部6は、一括して取得した複数の単語データがどの概念の下で同一のグループ内にグルーピングされているかを特定し、特定した概念が所定の概念に合致する場合にのみ、取得した単語データの一部または全部を処理に用いるものとしてもよい。

【0069】

また、エージェント処理部6は、トリガ取得処理、判別処理、入出力処理等の各種処理とワイヤとが全体として形成するフローを分担して行う、互いに接続された複数のデータ処理装置（例えば、コンピュータ等）から構成されていてもよい。この場合、解析処理部

3を構成するそれぞれのデータ処理装置は、解析処理部3が実行し得るフロー全体のうち、自己が実行する可能性がある部分を表すデータを、処理項目データベースやワイヤデータベースの要素として記憶すれば十分である。そして、それぞれのデータ処理装置が記憶するデータが、当該データ処理装置が実行する部分の処理をマクロ定義するようなデータとなっていれば、複数のデータ処理装置に分散処理を行わせることも容易である。

【0070】

また、このカーオーディオシステムは、音声入力部1や言語解析部2あるいは音声出力部4も複数備えていてよい。

また、音声入力部1は、たとえば、音声を表すデータが記録された記録媒体（たとえば、フロッピー（登録商標）ディスクや、CDや、MO（Magneto-Optical Disk）など）から波形信号を読み出して言語解析部2に供給する記録媒体ドライブ装置（たとえば、フロッピー（登録商標）ディスクドライブや、CD-ROMドライブや、MOドライブなど）を備えていてもよい。

【0071】

言語解析部2は、音声入力部1より供給された音声データが表す音声の話者（あるいは話者の性別など）を任意の手法で特定し、特定した話者を示す話者データをエージェント処理部3に供給するようにしてもよい。一方、エージェント処理部3は、例えば楽曲の再生を指示するたびに、楽曲データベースに、再生の対象となった楽曲が、話者データが示す話者のために再生されたことを示す履歴データを追加するようにしてもよい（あるいは、当該楽曲が当該話者のために再生された回数を示すデータを履歴データとして記憶しておき、この履歴データを1だけインクリメントするようにしてもよい）。そして、再生する楽曲を選択する場合は、このような履歴データに基づいて、楽曲を選択するようにしてもよい。

【0072】

以上、この発明の実施の形態を説明したが、この発明にかかるオーディオ機器制御装置は、専用のシステムによらず、通常のコンピュータシステムを用いて実現可能である。

例えば、入出力対象機器群5及び通信制御装置に接続されたパーソナルコンピュータに上述の音声入力部1、言語解析部2、音声合成処理部3、音声出力部4及びエージェント処理部6の動作を実行させるためのプログラムを格納した記録媒体から該プログラムをインストールすることにより、上述の処理を実行するカーオーディオシステムを構成することができる。そして、このプログラムを実行するパーソナルコンピュータが、図1のカーオーディオシステムの動作に相当する処理として、例えば、図4に示すフローを実行するものとする。

【0073】

なお、パーソナルコンピュータに上述のカーオーディオシステムの機能を行わせるプログラムは、例えば、通信回線の掲示板（BBS）にアップロードし、これを通信回線を介して配信してもよく、また、このプログラムを表す信号により搬送波を変調し、得られた変調波を伝送し、この変調波を受信した装置が変調波を復調してこのプログラムを復元するようにしてもよい。そして、このプログラムを起動し、OSの制御下に、他のアプリケーションプログラムと同様に実行することにより、上述の処理を実行することができる。

【0074】

なお、OSが処理の一部を分担する場合、あるいは、OSが本願発明の1つの構成要素の一部を構成するような場合には、記録媒体には、その部分を除いたプログラムを格納してもよい。この場合も、この発明では、その記録媒体には、コンピュータが実行する各機能又はステップを実行するためのプログラムが格納されているものとする。

【図面の簡単な説明】

【0075】

【図1】この発明の実施の形態に係るカーオーディオシステムを示す図である。

【図2】グルーピング用のフラグの具体例を模式的に示す図である。

【図3】ワイヤを示す図である。

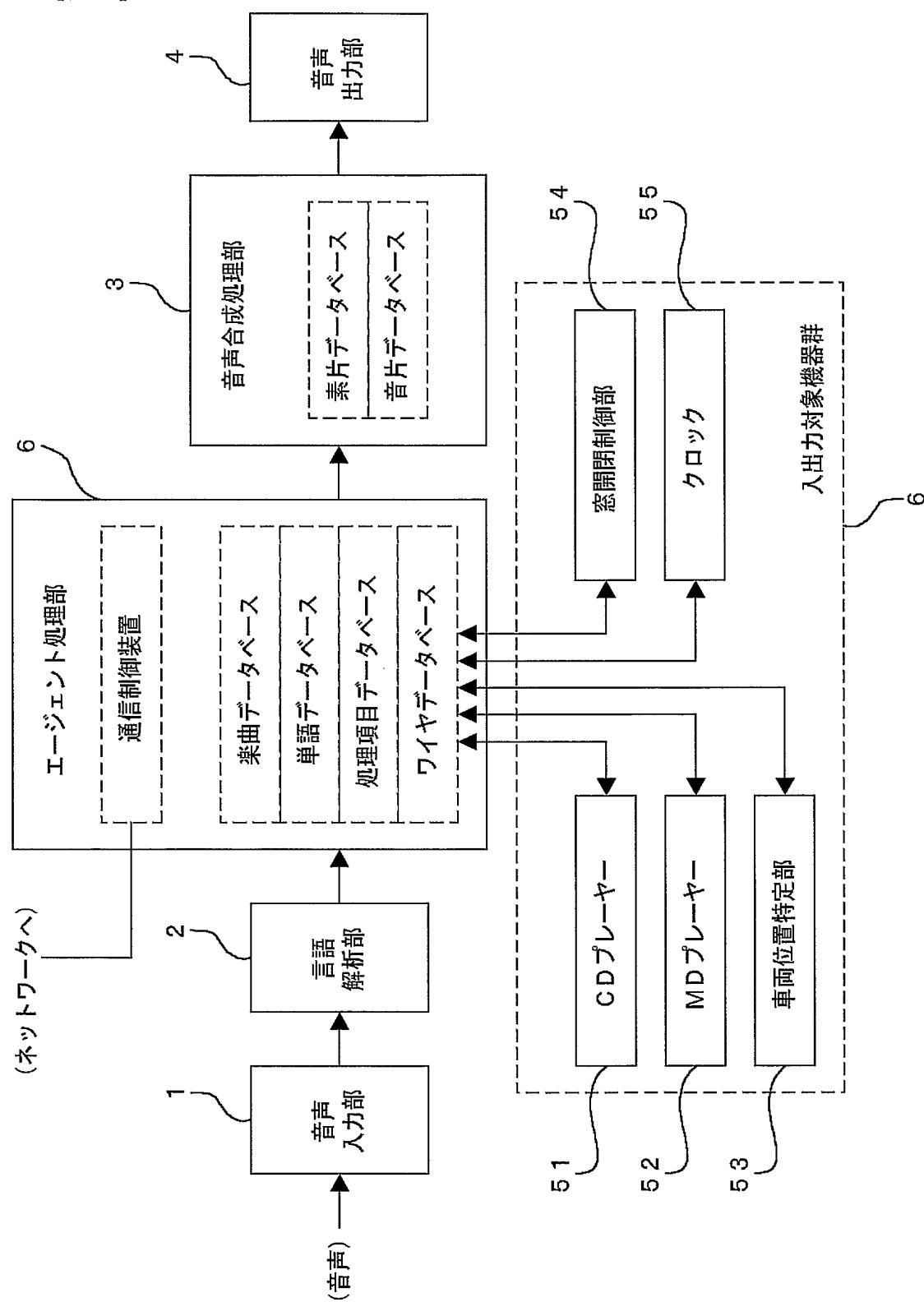
【図4】重み係数の設定を説明するための図である。

【符号の説明】

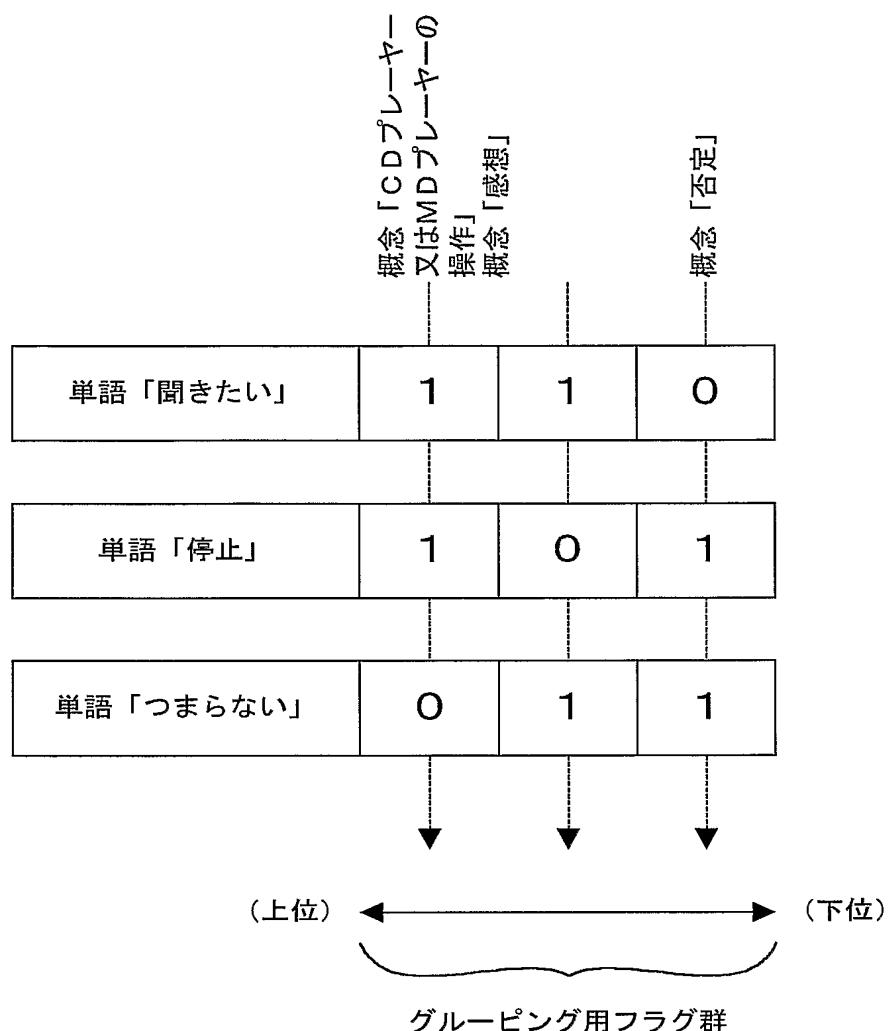
【0076】

1	音声入力部
2	言語解析部
3	音声合成処理部
4	音声出力部
5	入出力対象機器群
5 1	CDプレーヤ
5 2	MDプレーヤー
5 3	車両位置特定部
5 4	窓開閉制御部
5 5	クロック
6	エージェント処理部

【書類名】 図面
【図1】



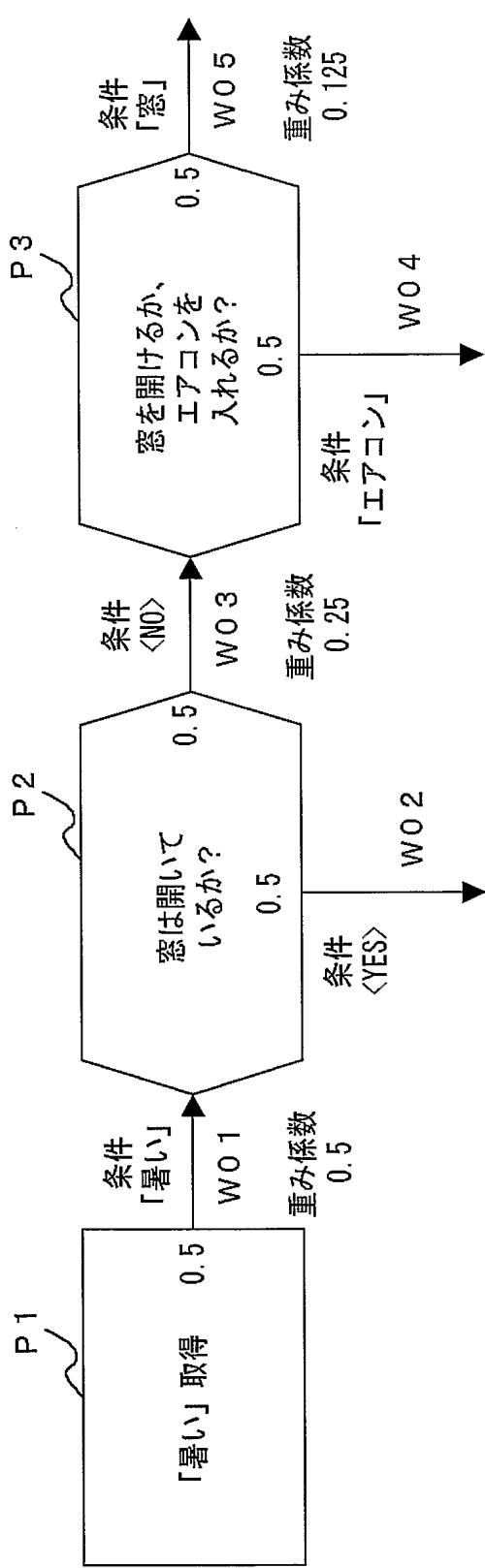
【図 2】



【図3】

```
W1 = From (.1) To (CN01.1), 1.0
W2 = From (CN01.3) To (EX01.1), 0.5
W3 = From (CN01.2) To (QB01.1), 0.5
W4 = From (.2) To (EX03.1), 1.0
W5 = From (QB01.3) To (EX03.1), 0.5
W6 = From (QB01.2) To (EX02.1), 0.5
W7 = From (EX02.2) To (.3), 0.5
```

【図4】



【書類名】要約書

【要約】

【課題】 人間が言語の形で発する指示に適切に応答して機器を制御できるオーディオ機器制御装置等提供することである。

【解決手段】 言語解析部2は、音声入力部1が入力した音声に音声認識を施して、この音声が表している可能性のある単語とそのスコアとを特定し、これらを表す単語データをエージェント処理部6に供給する。エージェント処理部6は、単語データ等を取得するデータ取得処理や、判別処理、入出力処理を定義する処理項目データと、ひとつの処理から次の処理への遷移を定義しこの遷移に重み係数を与えるデータであるワイヤとを記憶しており、処理項目データとワイヤとが全体として表すフローを実行することにより、入出力対象機器群5に属する機器を制御する。フロー内のどの処理へと遷移するかは、処理が進んだ地点とワイヤの接続関係とにより決まる各ワイヤの重み係数と、単語データのスコアとにより決まる。

【選択図】 図1

特願 2003-406669

出願人履歴情報

識別番号 [000003595]

1. 変更年月日 2002年 7月26日

[変更理由] 住所変更

住所 東京都八王子市石川町2967番地3
氏名 株式会社ケンウッド